# Computational Geometry: Approximate Nearest Neighbor Search in Quadtrees

Don Sheehy

April 29, 2010

## 1 Shifting Quadtrees

Let $P$ be the input set of $n$ points in $\mathbb{R}^d$. A shift is just a translation of the plane by some vector $v$. Let $s > d$ be an odd integer that will denote the number of shifts. The $i$th shift, $v_i$ is defined to be $(\frac{i}{s}, \ldots, \frac{i}{s})$, for $i \in \{0, \ldots, s-1\}$. Let $z(x)$ denote the point $x$ with its bits shuffled. That is, $z(x) \in [0,1]$ is the representation of $x$ on the Z-order space-filling curve.

**Definition 1.1.** *A point $a$ is* central *in an interval $[B_0, B_1]$ of length $b$ if $a + \frac{b}{2s} = qb + r$ for some $q \in \mathbb{Z}$ and $b/s \leq r < b$. We say that $a$ is central in a box $B = [B_0^{(1)}, B_1^{(1)}] \times \cdots \times [B_0^{(d)}, B_1^{(d)}]$ if it is central in $[B_0^{(i)}, B_1^{(i)}]$ for every $i = 1 \ldots d$.*

**Lemma 1.1.** *If $B$ is a quadtree cell containing $q$ and some point $p \in P$, then either* $\mathrm{pred}(q)$ *or* $\mathrm{succ}(q)$ *is in $B$.*

*Proof.* All points of $B$ have the same leading $k$ bits in their binary representation of each coordinate for some $k$. Suppose for contradiction that some points $x, y \in P$ are $\mathrm{pred}(q), \mathrm{succ}(q)$, respectively and do not lie in $B$. Then, $p$ has a longer prefix in common with $q$ than either of $x$ or $y$. Thus, either $z(x) < z(p) < z(q)$ or $z(q) < z(p) < z(y)$. Thus, either $x \neq \mathrm{pred}(q)$ or $y \neq \mathrm{succ}(q)$, a contradiction. □

**Lemma 1.2.** *If $a$ is central in a cell $B$ at scale $b$ and $|a - p| \leq \frac{b}{2s}$ then $p \in B$.*

*Proof.* This follows straight from the definition of *central*. □

**Lemma 1.3.** *Let $p \in P$ be the nearest neighbor of $q$. If $p, q \in B$ at scale $b \leq k|p - q|$ and $q$ is central in $B$, then either* $\mathrm{pred}(q)$ *or* $\mathrm{succ}(q)$ *is a $k\sqrt{d}$-approximate nearest neighbor of $q$.*

*Proof.* By Lemma 1.1, either $\mathrm{pred}(q)$ or $\mathrm{succ}(q)$ is in $B$. Let $a$ denote the nearer of $\mathrm{pred}(q)$ or $\mathrm{succ}(q)$ to $q$. Since $a \in B$, each coordinate differs from $q$ by at most $b$. So, the total distance from $a$ to $q$ is bounded:

$$|a - q| \leq b\sqrt{d} \leq k\sqrt{d}|p - q| = k\sqrt{d}|NN(q) - q|.$$

So, $a$ is a $k\sqrt{d}$-ANN of $q$.

$\square$

**Lemma 1.4.** *For any $a \in [0,1]$ and any interval $B$ of length $b = \frac{1}{2^\ell}$, there is at most one shift $v_i$ such that $q$ is not central in $B$.*

*Proof.* Suppose for contradiction that there are two shifts $v_i, v_j$ such that

$$\left(a + \frac{i}{s} + \frac{b}{2s}\right) = q_i b + r_i, \text{ and} \tag{1}$$

$$\left(a + \frac{j}{s} + \frac{b}{2s}\right) = q_j b + r_j, \tag{2}$$

where $0 \le r_i, r_j \le \frac{b}{s}$.

These two equations imply that

$$\frac{1}{s}(i-j) - b(q_i - q_j) = r_i - r_j.$$

Multiplying both sides by $\frac{s}{b} = 2^\ell s$, we get

$$2^\ell(i-j) - s(q_i - q_j) = \frac{s}{b}(r_i - r_j).$$

By our supposition, the RHS is nonnegative and less than 1. The LHS is an integer. Thus, the RHS must be 0. So, we can rewrite it as

$$2^\ell(i-j) = s(q_i - q_j).$$

We chose $s$ to be odd so it does not divide $2^\ell$. So, $s$ divides $i-j$. However, $0 \le i, j < s$, so it must be that $i - j = 0$ and thus $i = j$, a contradiction.

$\square$

**Theorem 1.1.** *For some shift $v_i$, either* $\mathrm{pred}(q)$ *or* $\mathrm{succ}(q)$ *is a $O(d^{\frac{3}{2}})$-approximate nearest neighbor of $q$.*

*Proof.* Let $p$ be the nearest neighbor of $q$. By Lemma 1.4, there exists $v_i$ such that $q$ is central at the smallest scale $b \ge 2s|p-q|$. Let $B$ be the corresponding cell containing $q$ at scale $b$.

The choice of $b$ implies that $|p - q| \le \frac{b}{2s}$. Lemma 1.2 implies that $p \in B$.

Let $a$ be the nearer of $\mathrm{pred}(q)$ and $\mathrm{succ}(q)$ for shift $v_i$. Lemma 1.3 implies that $a$ is a $2s\sqrt{d}$-ANN of $q$. Since $s = O(d)$, we are done.

$\square$